Biostat/Epi 536 2024

HW 4 (9 questions on two pages). Use the data “esophcts” for this assignment

Items marked *optional* do not have to be submitted to obtain full credit for this assignment. Optional items are valuable opportunities to learn, but made optional because they are less useful as midterm preparation.

We will cover splines in class on 10/22/24, which are needed for Q6. Before then, you should be able to complete Q1- Q5, Q9(a)-(c), and get started on Q7-Q8.

1. Fit a logistic model relating binary cider exposure (> 10 g/day) to the risk of esophageal cancer, using grouped-linear adjustment for age and the six age groups in the variable agegp. Report the odds ratio and 95% confidence interval for binary cider exposure. *Optional*: Write a sentence appropriate for the results section of a scientific paper reporting the results of this analysis. Note, here and below: your sentence should include both a point estimate and CI for the parameter(s) of interest.

2. Fit a logistic model relating binary cider exposure (> 10 g/day) to the risk of esophageal cancer using indicator variables to adjust for age and the six age groups in the variable agegp. Report the odds ratio and 95% confidence interval for binary cider exposure. *Optional*: Write a sentence appropriate for the results section of a scientific paper reporting the results of this analysis.

3. (a) Compare the results for the exposure variable in the Q1 and Q2 analyses. Are results similar or very different?

(b) (*Optional*.) Which result would you prefer to report in a scientific article, and why?

4. Fit a logistic model relating binary cider exposure (> 10 g/day) to the risk of esophageal cancer using linear adjustment for age (as a continuous variable). Report the odds ratio and 95% confidence interval for binary cider exposure. *Optional*: Write a sentence appropriate for the results section of a scientific paper reporting the results of this analysis.

5. Fit a logistic model relating binary cider exposure (> 10 g/day) to the risk of esophageal cancer using quadratic adjustment for age (as a continuous variable). Report the odds ratio and 95% confidence interval for binary cider exposure.

6. (a) Fit a logistic model relating binary cider exposure (> 10 g/day) to the risk of esophageal cancer using linear spline adjustment for age. Use the same age groups as for the ageg variable for your splines. Report the odds ratio and 95% confidence interval for binary cider exposure.

(b) (*Optional.*) Write 1-2 sentences describing the analysis (not the results) that would be appropriate for the methods section of a scientific paper.

7.

(a) Compare the OR estimates from Q4, Q5, and Q6. Are your results similar or very different?

(b) (Optional.) Which approach would you prefer if you were studying this exposure and wanted to adjust for age as a potential confounder, and why?

8. For each model in Q1 through Q6, how many degrees of freedom are used to include age in the model? That is, how many model parameters are used for age? Include the model intercept in your count.

9.

(a) Are the models in Q1 and Q2 nested?

(b) Are the models in Q1 and Q4 nested?

(c) Are the models in Q4 and Q5 nested?

(d) Are the models in Q4 and Q6 nested?

(e) Are the modest in Q5 and Q6 nested?

For each question (a)-(e), if you say that models are nested, write out the full and “reduced” models, and give the restrictions on the full model that yield the reduced model.